Foot Contact Timings and Step Length for Sprint Training

Murray Evans Steffi Colyer Darren Cosker Aki Salo
CAMERA centre, University of Bath
m.evans@bath.ac.uk

Abstract

The frequency and length of a runner's steps are fundamental aspects of their performance. Accurate measurement of these parameters can provide valuable feedback to coaching staff, particularly if regular measurement can be made and monitored over the course of a season. This paper presents a computer vision based approach using high framerate cameras to measure the location and timing of foot contacts from which step length and frequency can be determined. The approach is evaluated against force-plates and optical motion capture for a mix of 18 trained and recreational runners. Force-plates and optical motion capture are considered to be the current "gold-standard" in biomechanics, and this is the first vision based paper to evaluate against these standards. Landing and take-off times were shown to be measurable to within 1.5 frames (at 180fps) and step length to within 1 cm.

1. Introduction

The use of video to analyse the technique and performance of athletes is an established area, from the manual annotation of videos [12] through to full-body motion capture using optical marker-based tracking systems [18] and emerging research systems based on markerless technologies [9]. A key aspect of the application of video analysis to biomechanics and sport is the need for high precision in both time and location (whereas a graphics application, more typical to the computer vision literature, can get by with only accurate appearance).

Running performance is affected by many factors, but two fundamental properties affecting speed are step length and step frequency, with duration of contact also of interest. Step frequency and contact start and end times can be measured to high precision using force plates embedded in the floor of a running track [10], while step-length can be reliably estimated [8] from optical marker-based motion capture systems (such as Qualisys and Vicon). Force-plate systems are expensive and not widely available, with very few facilities equiped to handle more than a couple of steps. Optical motion capture systems can be made available at tracks, but are again expensive, and more often located in dedicated labs and studios. The primary problem with marker-based motion capture systems is the requirement for the athlete to wear a set of infra-red (IR) reflective tracking markers. These markers, which must be accurately positioned on the body, can interfere with the natural performance of the athlete, and can be time consuming to position and thus impractical for regular use.

Inertial measurement units [15] can be made regularly available at training facilities, however these also require the athlete to wear equipment which may interfere with their natural performance. Optojump, a commercial system based on the athlete breaking light beams with their feet appears to provide a solution that does not impact performance, but the accuracy of foot contact detection has been questioned [3]. Currently the most appropriate method of obtaining step characteristic information outside of the laboratory, without impeding the athlete, is to use manual digitisation of video sequences [4]. However, this method is laborious and time-consuming, which limits its utility.

Automatic video based detection techniques can provide step-length and contact-time measurements non-invasively, but achieving high accuracy is difficult. One approach is to observe that the runner's foot is effectively static once on the ground (though the foot does still rotate around the toes). This provides a strong cue from which ground contact can be determined. For example, Zhu *et al*. [24] observed that motion blur reduced the clarity of edges on the moving foot vs. when the foot was static, allowing them to measure step length and frequency. However, such an approach seems sensitive to camera shutter speed and this was not addressed. A more promising approach is to use the silhouette of the runner (from background subtraction) to compute an "accumulator" image. Initialised to 0, the accumulator then counts up the number of frames during which an image pixel is labelled foreground. In Harle *et al*. [11] this technique was used with the runner passing horizontally across the image, and each step's contact identified by applying a threshold to the accumulator. From this they could measure step frequency and step length (in image space). The

basic algorithm then required extra work to eliminate false-positive contacts, but shadows, reflections or sub-optimal running technique could all still affect performance. The accumulator approach was also used by Jung and Nixon [13] to localise foot contacts, but step timings were indirectly inferred from the periodic movement of the person's torso. Another approach based on indirect observations was taken by Amini *et al*. [2], where an RGB-D camera (Microsoft's Kinect v2) was used to track knee angle, and this used to infer foot contact events. Performance was reported for detection accuracy, but not timing or step length precision. The idea of using RGB-D sensors could have advantages for body and foot localisation, however the relatively slow 30 Hz sampling rate of most such systems limits applicability in fast-motion, timing critical measurement systems. Similar indirect measurements could be made through modern, markerless motion-capture techniques such as OpenPose [7], but is not clear that suitable precision could yet be achieved.

Good precision in measuring step-length and contact times of individual steps would allow coaches to build up statistics of their athletes during multiple stages of events, such as the start, the acceleration phase, and the maintenance of peak velocity. Where errors are too large, averaging is required over more steps to produce robust performance statistics, reducing the potential information gains for coaches. As such, this paper focusses on measuring the timings of individual contacts, rather than a broader step frequency.

A multi-camera system is proposed. Although the use of multiple cameras increases the expense and difficulty of hardware set up, the added robustness is considered a significant advantage. The proposed system is the first such vision system to be evaluated against gold-standard biomechanics measurement devices: a force-sensitive plate embedded in the floor beneath the runners for timings, and marker based motion capture of a subset of the runs for step length. Results show that contact start and end can be timed to within 1.5 frames (at 180fps), and step length to within 1 cm.

2. Proposed approach

2.1. System overview

The proposed approach uses multiple synchronised and calibrated cameras to perform robust and accurate localisation of each foot contact. The determination of the exact frame when the foot makes contact with and subsequently takes-off from the ground is determined using a single camera algorithm. The overall structure of the system can be seen in Figure 1.

The arrangement of cameras is designed such that an extended number of cameras could be used to form a "corridor" lining a running track, as shown in Figure 2. The main cameras are arranged with optical axes perpendicular to the running track to provide the desired view for determining contact times, and so that any foot contact within the corridor should be observed by three cameras. Extra cameras with a view more parallel to the track are used to aid localisation. The perpendicular cameras should be sufficient on their own, but in the practice the extra track-parallel cameras have been vital to ensuring robust contact detection, making the 5-camera setup a minimum setup. The cameras have wide angle lenses to capture multiple steps from each section of the corridor, and to facilitate the imagery also being used for whole-body motion analysis should a suitable parallel system be available.

The system requires synchronised and calibrated cam-

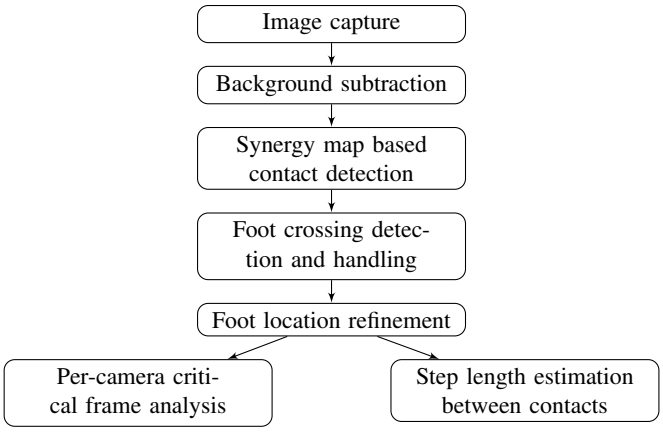

Figure 1. Overall system

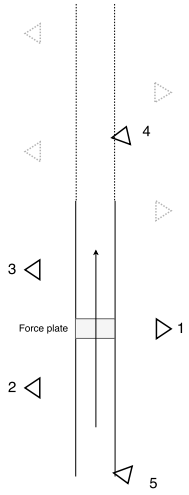

Figure 2. Cameras 1, 2 and 3 form a triangle observing a section of track. An extended corridor can be created by duplicating this triangular arrangement (dotted triangles). Individual foot-contacts should thus be viewed by three cameras. Cameras 4 and 5 can aide localisation by viewing more parallel to the direction of travel. The force plate serves as a validation tool.

eras. Calibration is achieved through observations of a calibration board. Instrinsic calibration is achieved as per Zhang [23]. The known size of the calibration board allows each board position to be reconstructed in 3D from a single camera. One camera is chosen as the "root" and set at the origin, and the boards it observes are initialised in 3D. Cameras that can also observe those board positions can then be initialised relative to the boards. The new cameras allow more boards to be initialised, allowing more cameras to be initialised. Bundle adjustment [20] is used to optimise the positions of cameras and grids using the Ceres solver [1]. With practice, this calibration procedure is not difficult and the resulting calibration is suitable for precise foot localisation. The final stage of calibration uses manually annotated points on the ground to identify the scene origin and ground plane alignment. It is assumed that the ground is approximately planar, and the scene coordinate system is aligned such that the ground forms the $z = 0$ plane (referred to as the "ground plane"), with $+z$ up. For simplicity, the $+y$ axis was aligned with the direction of running, and the origin was at the corner of the force-plate.

2.2. Foot contact detection

A multi-camera approach is used for detecting the approximate time and precise location of each foot contact. First, background subtraction is performed to isolate the athlete in each video stream. Background subtraction is suitable for real-world outdoor use so long as a degree of environmental control is enacted to control shadows, reflections, sudden lighting changes, and sufficient contrast between the athlete and the background. Contrast is especially important regarding the athlete's footwear and the track. A recent and comprehensive review of background subtraction approaches is provided by Bouwmans [6]. For the results presented in this paper, the IMBS algorithm [5] was used.

For each frame t of video, the image from camera i is denoted as $I_{i,t}$, and the foreground mask resulting from background subtraction denoted as $M_{i,t}$. The foreground mask labels each pixel of the image as background L_b, shadow L_s or foreground L_f, as shown in Figure 3. Ideally, the runner should be segmented wholly as foreground, with all other pixels labelled as background or shadow. In practice, the settings of the background subtraction that allow for complete body segmentation also result in some image noise being labelled as foreground. Such noise regions will normally be inconsistent between cameras, and multi-camera processing will allow them to be handled.

2.2.1 Ground occupancy map

Foot contact events and their approximate locations are determined using a "synergy map" style approach [14]. A grid G is defined over the ground plane, with n_c columns and n_r rows - this is the "synergy" or "ground occupancy" map. The centre $p_g(r,c) = [\ x_{r,c} \quad y_{r,c} \quad 0\ ]^T$ of each cell $g(r,c)$ of this grid can be projected, using the camera calibrations, into each camera view, as $(u_i, v_i) = P_i(p_g(r,c))$ (where P_i is some function representing the projection). This computation need only be computed once to improve processing efficiency.

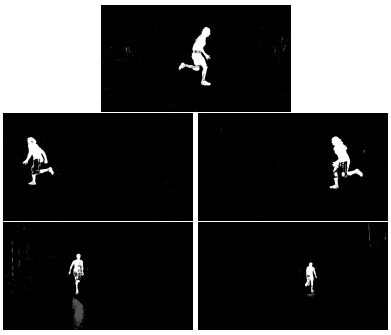

Figure 3. Background subtraction results for a frame of video of a runner. Foreground is shown in white, shadows are grey, background is black.

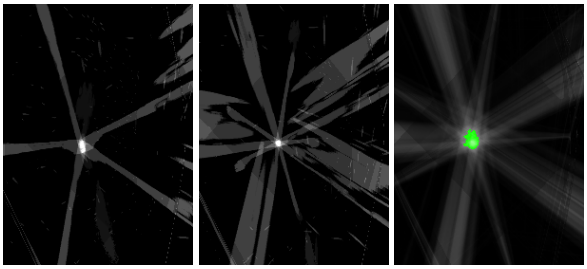

Figure 4. From left to right: Ground-plane occupancy before thresholding, knee-plane occupancy before thresholding, and body occupancy with green highlight where values are larger than threshold. These can be thought of as top-down views of the scene. The brighter a pixel is, the more cameras see that point of the scene as foreground. These occupancy maps correspond to the foreground masks in Figure 3

The synergy map for each frame of video is constructed by first initialising each grid cell to 0. Each cell is then processed to count:

- the number $v_{r,c}$ of camera views the cell is visible in (the cell's centre projects within the image)
- the number $o_{r,c}$ of camera views that see the cell as foreground (*i.e.* $M_{i,t}(u_i, v_i) = L_f$).

When the athlete runs through the frame, their feet will contact the ground, and their location on the ground will be identifiable through the occupied cells for which $\frac{o_{r,c}}{v_{r,c}} = 1$. This can be seen in Figure 4.

The ground occupancy map can be computed for every frame of the video. Next, for each cell, the set of *activation periods* is determined. An activation period is defined

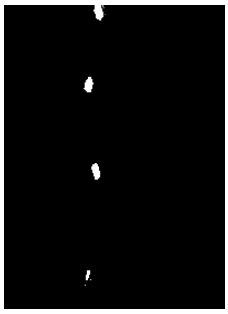

Figure 5. Image depicting ground cells for which the longest activation period has a duration longer than a specified threshold. Individual foot contact regions are clearly visible.

as an unbroken set of frames for which a cell is occupied, and recorded with its start and end frames. Activations with shorter duration than a specified threshold are discarded as noise.

2.2.2 Body occupancy map

It is possible for noise in the foreground masks, or other distractions, to cause spurious occupancy of grid cells. To help improve robustness, a process is implemented that determines the location of the athlete's torso, and verifies that only activations that conform with the location of the body can be considered as potential foot contact locations.

The body occupancy map B is constructed based on a modified synergy map approach [22] that has the effect of merging multiple horizontal scene planes. For each grid cell, a line is drawn vertically (parallel to the scene z axis) with its end points at $[\ x_{r,c} \quad y_{r,c} \quad 0.75\ ]$ and $[\ x_{r,c} \quad y_{r,c} \quad 2.0\ ]$. These end points are projected into each image, and the percentage of pixels along the resulting line segment in the image that are foreground is calculated, accumulated across views, and normalised by the number of views. The resulting value is thresholded, and cells with a value larger than the threshold are labelled as occupied. An example of a body occupancy map can be seen in Figure 4.

2.2.3 Identifying foot contacts

The easiest method for identifying foot contacts is to create an image where each cell is coloured based on the length of its longest activation period. As shown in Figure 5, this allows for a simple extraction of contact locations by thresholding on activation duration. However, such an approach is not robust to multiple runners or a runner returning over the same ground. Although these events are not expected in the basic use case of a single runner passing through the scene, a more flexible algorithm is still preferred.

Frames of the video are processed iteratively. At each frame, all grid cells are explored and those which have an activation period including the current frame are identified. Some extra work is performed to better handle noise on the foreground mask which can briefly cause some occupied cells to appear not-occupied. The earliest start frame t_e and latest end frame t_l from all activation periods that include the current frame are identified. Any unoccupied cells that have activation periods ending after t_e or starting before t_l are also considered occupied in the current frame.

A binary image where each pixel represents the cells of G as either occupied or empty is created. A similar image is constructed for cells in B which are occupied during the time window from $t_e \rightarrow t_l$. The two grid occupancy images are intersected, and the result is dilated to help fill any small holes. Connected components is then used to identify the blobs in this image as a set of potential foot contacts active at this frame.

The potential foot contacts are either recorded as new contacts (regions of grid cells with a known start and end time), or used to update contacts at the same location identified in the previous frame. Once all frames of video have been processed, a set of foot contacts will be known, as well as the frames at which they start and end. An approximate position for the contact is taken as the centre of mass of occupied cells on the frame at which the contact appears largest. Multiple contacts in the same location at the same time are merged together on the assumption that only one foot will be present at any one location at any one moment.

2.3. Contact location refinement

The initial location of the contact is imprecise and generally needs to be improved. A 3D bounding box with dimensions approximately foot-sized (380mm long, 60mm wide, 35mm tall) serves as a basic model for the foot. It is aligned with the direction of motion of the runner, and initialised with its centre at the initial estimate (x_0, y_0) of the foot contact position.

The estimate of the location of the foot is refined by evaluating the position of the bounding box, and minimising a suitable error function. The proposed error function consists of four terms:

$$\epsilon(x,y) = w_d\epsilon_d(x,y) + \sum_i (w_f\epsilon_f(x,y,i) + w_s\epsilon_s(x,y,i) + w_c\epsilon_c(x,y,i)) \quad (1)$$

The first error is a simple distance from the initialisation position. This helps to ensure the foot contact estimate does not get distracted by other foreground regions in the image and remains relatively close to the initial foot position.

$$\epsilon_d(x,y) = \left\| \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \right\|$$

The remaining three error terms are summed across all cameras i. Each requires the projection of the 3D bounding

box β into a 2D bounding box β_i in each image. This is achieved by projecting the 8 corner points of β into image i and finding the smallest axis-aligned bounding box that encapsulates all of the projected corners.

The first of the remaining three parts of the error, $\epsilon_f(x, y, i)$ is the sum of foreground pixels within β_i, normalised by the area of β_i. Note that this error is negative because the aim is to find a *minimum* of the overall error term. The normalisation ensures that a solution with a larger bounding box (by being closer to a camera) is not preferred to one with a smaller bounding box:

$$\epsilon_f(x, y, i) = -\frac{1}{A(\beta_i)} \sum_{p \in \beta i} k(p) \tag{2}$$

$$k(p) = \begin{cases} 1 & \text{if } p = L_f \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

For the next term, the foreground mask image of each view M_i is transformed into a distance image D_i, where the value of each pixel represents the distance of that pixel from the nearest foreground pixel (see Figure 6). By summing the value of these pixels inside β_i, this helps to direct the search for a solution towards foreground regions.

$$\epsilon_s(x, y, i) = \sum_{p \in \beta_i} D_i(p) \tag{4}$$

The final term is designed to encourage the bounding box to centre on the foot. By being larger than the foot, there can be multiple bounding box positions that wholly enclose the foot, and thus have an equal number of foreground pixels. This term helps to reinforce the desire for the foot to be centred in the bounding box. In each view, the leftmost $p_l = (u_l, v_l)$ and rightmost $p_r = (u_r, v_r)$ pixels inside β_i which are foreground are found, as well as the left $l_{\beta,i}$ and right $r_{\beta,i}$ edges of β_i. The error term tries to ensure the distance from the left edge of the bounding box to the leftmost foreground pixel is equal to the distance between right edge and rightmost pixel.

$$\epsilon_c(x, y, i) = |(|u_l - l_{\beta,i}|) - (|u_r - r_{\beta,i}|)| \tag{5}$$

The overall error can be minimised using standard optimisation algorithms such the Nelder-Meade Simplex [19]. The weights have been set to $w_d = 500$ (assuming distances are measured in millimetres), $w_f = w_s = 1$ and $w_c = 2000$. The search will optimise only the x and y position of the bounding box centre, with z set such that the base of the bounding box is on the ground, i.e. $[x \quad y \quad 17]^T$ Orientation of the bounding box is currently assumed to be known from the running direction.

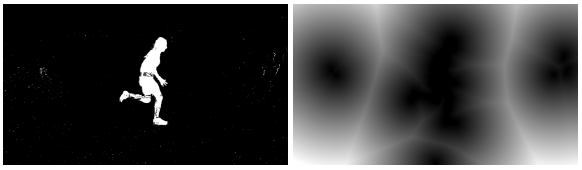

Figure 6. Distance transform (right) of the foreground mask (left)

2.3.1 Selecting a frame for refining contact location

For many runners, especially recreational runners, the static foot making contact with the ground can be obscured by the second foot as it crosses through its step. To get the best estimate of the location of the static foot, it is important to avoid moments when the crossing foot causes any occlusions or interference.

To this end, it is useful to construct a further occupancy map at approximately knee-height. This can be used to show the path of the crossing foot and identify moments when the crossing foot does not interfere with observations of the static foot. The occupancy map of the knee plane can be imaged such that each occupied cell takes on a brightness from 0 (the earliest estimated start of the contact) to 1 (the estimated end time of the contact), based on the last frame the cell is considered occupied. This will produce an image with a bright patch where the static foot is (as it remains present until the end of the contact), and a streak through the image where the crossing foot moves through. Examples of this can be seen in Figure 7.

After identifying where in this map the static foot is located, the time at which the crossing foot is one foot-length in front of and one foot-length behind the static foot can be determined. If the crossing foot is not present behind the static foot, or more rarely, not present in-front of the static foot, then the crossing foot is unlikely to cause interference with observations of the static foot. However, if the presence of the crossing foot is detected, the time where it is one foot-length *behind* the static foot is chosen as the moment for refining the position of the static foot.

2.4. Critical landing and takeoff frames

A precise estimate of step-frequency requires a precise estimate of the exact moment the foot makes contact with or takes-off from the ground. Contact duration is measured from the frame the foot first makes contact, to the frame after it is last in contact. These instants can be difficult to observe, leading to indirect approaches for finding step timing using knee [2] or ankle joint angles, or head motion [13]. Indirect inference could also be made by observation of the periodic motion of the foot. However, as with other indirect observations, the precise contact times are unclear.

Observation of the foot on each contact suggests that a direct observation of contact start and end can be determined by watching for the start/end of vertical motion of the

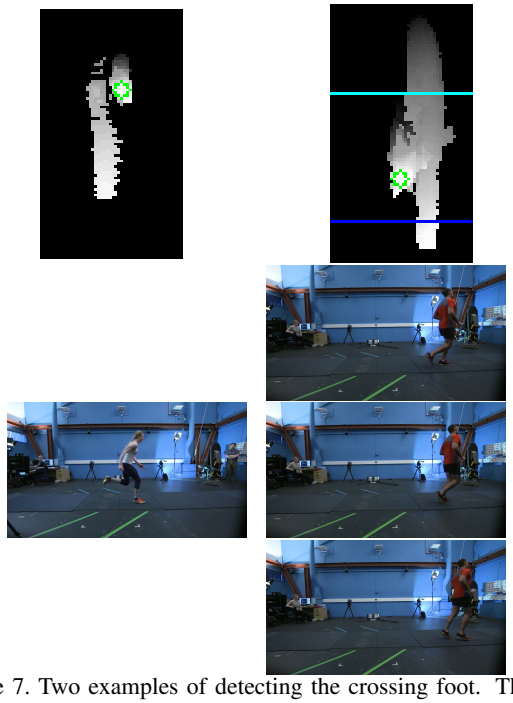

Figure 7. Two examples of detecting the crossing foot. The top shows a section of the knee occupancy map, with shading representing time from contact start to contact end, and the green circle the location of the static foot. On the left, the trained sprinter's technique with high knees causes no crossing issue. On the right, the motion of the crossing-leg through the knee plane can be seen. The crossing event is detected, and the lines across the image show one foot-length in front of and behind the static foot. These lines allow the determination of a time when the crossing foot does not obscure the static foot.

foot. However, as the foot is a deformable object, tracking it as a single entity (for example using algorithms submitted to the Visual Object Tracking Challenge [16]) will not robustly identify when the first part or last part of the foot makes contact with the ground. As such, tracking individual parts of the foot is more informative. Traditional feature points trackers such as SIFT [17] were found to have difficulty locking onto specific small parts of the foot, particularly with sufficient density. As such, and with respect to the goal of monitoring purely vertical movement, a task specific *slice* image feature was developed.

2.4.1 Slice features

The slice image features are constructed as follows. The bounding box β_i (see Section 2.3) is used to isolate the foot in the image. The height of β_i is tripled by padding above and below. A subwindow W of the video frame is now extracted. W is lightly smoothed using a bilateral filter to give W_s, and then vertical gradients W_g of W_s are calculated. W_s and W_g are subdivided into n_s vertical slices. Where the slices are more than one-pixel wide, the horizontal mean of pixel values is calculated such that each vertical slice becomes a 1-D vertical feature. These colour S_c and gradient S_g slices (Figure 8) are used as tracking features. For the results presented in this paper, 30 slices were used.

Figure 8. Top row: Colour image and vertical gradients. Bottom row: Slice features assuming image divided into 10 slices (30 slices are used for the presented results).

2.4.2 Tracking slices

Tracking always begins from a frame when the foot is known to be on the ground. For take-off, tracking is performed forwards in time. For landing, tracking is performed backwards in time. To avoid confusion with the crossing foot, tracking is started using frames where the crossing foot will not be a problem. These frames are identified during pre-processing for foot location refinement as described in Section 2.3.

Tracking begins at the identified start frame and proceeds for a duration of 0.2 seconds. This time is set as a configuration option and is long enough for both trained runners and recreational runners to complete each contact.

Colour and gradient slice features are computed for all tracking frames. Let the slices in frame t be denoted as S_t, and let s_{tn} be slice number n in S_t, where s_{tn} consists of the colour slice c_{tn} and gradient slice g_{tn}. Further, let $s'_{tn} = c'_{tn}, g'_{tn}$ denote a cropped version of s_{tn} corresponding to the unpadded bounding box β_i (i.e. the central third of s_{tn}) with n_r rows.

An evidence matrix $M(o,n)$ is now computed for vertical offsets o and slices n:

$$M(o,n) = \sum_{r=0}^{r_n} \|c'_{tn}(r) - c_{(t+1)n}(r+o)\| + \sum_{r=0}^{r_n} \|g'_{tn}(r) - g_{(t+1)n}(r+o)\| \quad (6)$$

Individual slices could be tracked by taking $\arg\min_o M(o,n)$. However this can be prone to noise, giving the impression that the slice has moved when it has

not. To combat this, belief propagation [21] is used to share information between neighbouring slices (with N the set of neighbours). For this, let O_s be the set of possible tracking offsets, and o_s be the offset assigned to slice s in the set of slices S_t. An energy function is defined:

$$E(\mathbf{O}) = \sum_{s \in S} M(o_s, s) + \sum_{m,n \in N} w_s V(o_n, o_m)$$

Here, w_s is a weight on $V(a, b)$, a smoothing function that considers the tracking offsets assigned to two neighbouring slices a and b, penalising them if they are different. This specific smoothing term is designed to balance the need for smoothness, with the need to allow a slice to remain static if its local information indicates as such.

$$V(a, b) = \frac{1}{(1 + e^{4-0.5(a-b)^2})}$$

Belief propagation solves for $\arg\min_O E(O)$, giving the optimal tracking offset to apply for each slice. It is important to control the weight of the smoothness. Set it too strong, and slices can move too early due to motion of their neighbours. Too weak, and tracking can be noisy, causing spurious motion of slices.

2.4.3 Deciding on critical frames

The critical frames on individual cameras are taken in each case to be the frame when the last slice that moves shows its first motion. Each camera makes an independent observation of the critical contact start and end frames, producing more than one result for each foot. As such, it is necessary to combine the independent camera results into a single answer.

The cameras observe the foot from different vantage points, slightly ahead, or slightly behind the foot. For trained sprinters, the foot lands on the fore-foot, with the heel possibly not making contact. The foot then rolls forwards around the fore-foot leaving the tip of the toe as the last contact point. A camera that is slightly ahead of the foot is most likely to have the best view of the first and last part of the foot to make contact (the toe in each case). A recreational runner has an increasing likelihood to land flat-footed, or heel first, and then roll through the toes.

A camera that is slightly behind the foot will tend to trigger early for take off (as the exact end of the toe is not visible), and for trained runners, will also trigger early (thus late - tracking is performed backwards in time) on landing. The result is to observe that it is not necessary to select the camera with the best view, but only to take the camera that reports the latest takeoff, and the earliest landing.

3. Performance analysis

3.1. Test setup

The vision system was configured with 5 cameras as shown in Figure 2. The cameras were Sony PXW-FS7 TV cameras, set to record HD-video (1920x1080) at 180fps and using wide 10 mm lenses. This specific setup positioned the cameras about 2m from the centre line of the track, with 3.5m between cameras 2 and 3, but exact camera positions will depend on specific installations and available space. At 180fps, these cameras could not be genlocked (a frame level sychronisation technology), so synchronisation was handled through observations of a set of timing lights. The timing lights were three separate, synchronised series of 20 LEDs (Wee Beasty Electronics, UK), at least one set visible to each camera.

Foot-ground contact timings computed by the proposed vision algorithm were compared against those obtained using the "gold-standard" biomechanical measurement techniques of force plates and marker-based motion capture. Two force platforms (Kistler, 9287BA, Kistler Instruments Ltd., Switzerland; 1000 Hz) embedded in the laboratory floor provided precise timings of touchdown and take-off, which were considered to occur when force increased above or decreased below 5 N, respectively. A 200 Hz Qualisys motion capture system with ten cameras was used to track markers on a subset of the runner's feet. The difference in the y position of the foot during mid-stance was computed to provide step length. Triggering of the timing lights also triggered the motion capture and force-plate systems, ensuring sychronisation of all systems.

Eight well-trained sprint athletes (athletics and bob skeleton) and 10 recreational runners participated in the validation study. Each participant performed a total of 10 runs across the laboratory. The sprint group wore spikes for five of these trials and normal running trainers for the remaining five. This was to assess for any influence of footwear on the performance of the algorithm. The recreational runners completed all 10 trials in normal trainers. For five of these trials, passive reflective markers were placed on the toe and the heel. This allowed step length to be calculated using the marker-based system and also allowed any influence of the markers on the vision system's performance to be evaluated. Typically, motion data for three complete steps were captured in each trial (137 comparisons in total for step length). Recreational runners generally contacted both force platforms (yielding two comparisons of the contact timings per trial), whereas the sprint athletes only made contact with one, resulting in a total of 263 comparisons for contact timings.

Differences (absolute and signed) between the foot-ground contact timings and step length calculated by the computer vision algorithm and the biomechanical measure-

error	abs	sig	std
recMarks	9.3	-0.3	12.7

Table 1. Mean absolute error (abs), mean signed error (sig) and std-deviation of signed errors, for step length in mm

ment system were computed for each trial. Averages were computed for each participant across all ten trials, and for each condition (different footwear and with/without markers).

3.2. Results

Results indicate performance for subsets of the runners, trained with spikes (tSpikes) or normal trainers (tTrains), recreational runners with (recMarks) and without (recNorm) markers.

3.2.1 Step Length

Table 1 summarises the results for step length errors, which are only available for the runs of recreational runners wearing markers (n = 137 steps). The 9.3 mm mean absolute length error is within the 1 cm target for sports biomechanics. The largest errors occured when the contact was at the extremes of the observed area, compromising camera visibility. An extended corridor of cameras would be expected to resolve this problem - guaranteeing each contact is seen from enough cameras.

3.2.2 Contact timing

Table 2 summarises the results for landing and takeoff frames, and for contact duration errors (80 contacts for trained sprinters and 183 recreational). In general, the type of footwear did not affect the timing accuracy.

Trained athletes, and to an extent, landings, were easier due to their faster motion. Runners with faster contacts tended to have faster vertical acceleration causing a more obvious frame when the foot was in the air vs. when it was on the ground. Contact durations varied from 23 ms to 28 ms for trained athletes, and 30 ms to 60 ms for recreational runners. Contact times were easier for the algorithm to detect when the observed motion was fast, as this gave a larger change between on-ground and off-ground frames. Slower runners suffered larger tracking errors, and in the extreme, very relaxed runners, on some steps scuffed/skimmed the floor rather than lifting. This lack of vertical motion effectively violates the fundamental assumption of the slice tracking algorithm, resulting in the largest timing errors.

These results are compared against the single-camera "accumulator" approach of Harle *et al.* [11]. The two approaches have comparable results for take-off of trained athletes. However as the runners get slower, the slice-tracking approach of this paper is consistently more precise. Landing

error		this paper			accumulator		
		l	t	d	l	t	d
tSpikes	abs	0.7	1.3	1.0	5.9	1.1	5.3
	sig	-0.4	-1.1	-0.7	5.9	0.6	-5.3
	std	0.8	0.9	1.1	2.1	1.1	2.4
tTrains	abs	0.8	1.2	1.0	4.6	1.6	4.9
	sig	-0.6	-0.9	-0.3	4.6	-0.2	-4.8
	std	0.8	1.1	1.3	2.5	1.8	2.4
recMarks	abs	1.8	1.9	3.2	5.2	3.9	3.7
	sig	-1.6	1.0	2.6	5.2	2.3	-2.9
	std	1.7	2.4	2.9	3.6	4.1	4.7
recNorm	abs	1.9	1.6	3.0	5.6	4.1	3.3
	sig	-1.8	1.1	2.9	5.4	2.9	-2.5
	std	2.0	1.5	2.4	3.3	3.7	3.6
all	abs	1.5	1.6	2.5	5.4	3.2	4.0
	sig	-1.3	0.4	1.8	5.3	1.9	-3.4
	std	1.7	2.0	2.8	3.2	3.6	3.9

Table 2. Mean absolute error (abs), mean signed error (sig) and std-deviation of signed errors, as number of frames at 180fps, for (l)anding, (t)akeoff and contact (d)uration.

is much more ambiguous with Harle's approach, due to deciding whether contact starts with the first pixel to pass the time threshold, or the last. The first tends to produce early contact while the last (used here) produces a late contact.

4. Conclusion

A multi-camera, markerless technique for measuring foot contact times and step-lengths for sprint athletes and runners has been presented. The method is compared against "gold-standard" techniques used in sports-biomechanics in the form of a force-plate embedded in the floor (for contact time) and optical marker-based motion capture (for step-length). The presented algorithm is capable of determining step-length to within 9 mm error on average, and contact times to 1.5 frames on average (at 180fps). This is more precise than a previously published single camera approach [11], and also good enough to be useful for sports-biomechanics applications. Future work will be undertaken to verify the performance in a wider range of environments, with a view to prototyping the technology for use by sprint teams.

5. Acknowledgements

This research was funded by CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications, EP/M023281/1.

References

[1] S. Agarwal, K. Mierle, and Others. Ceres solver. http://ceres-solver.org. 3

[2] A. Amini, K. Banitsas, and S. Hosseinzadeh. A new technique for foot-off and foot contact detection in a gait cycle

based on the knee joint angle using microsoft kinect v2. In *2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pages 153–156, Feb 2017. 2, 5

[3] R. Ammann, W. Taube, and T. Wyss. Accuracy of partwear inertial sensor and optojump optical measurement system for measuring ground contact time during running. *Journal of Strength and Conditioning Research*, 30:2057–2063, 2016. 1

[4] I. Bezodis, A. I. T. Salo, and D. Kerwin. A longitudinal case study of step characteristics in a world class sprint athlete. In *Proceedings of XXVI International Conference on Biomechanics in Sports*, pages 537–540, 2008. 1

[5] D. D. Bloisi, A. Pennisi, and L. Iocchi. Parallel multi-modal background modeling. *Pattern Recognition Letters*, 96(Supplement C):45 – 54, 2017. Scene Background Modeling and Initialization. 3

[6] T. Bouwmans. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11(Supplement C):31 – 66, 2014. 3

[7] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017. 2

[8] M. Coh, S. Peharec, P. Bacic, and K. Mackala. Biomechanical differences in the sprint start between faster and slower high-level sprinters. *Journal of Human Kinetics*, 56:29–38, 2017. 1

[9] A. El-Sallam, M. Bennamoun, K. Honda, A. Lyttle, and J. Alderson. Towards a fully automatic markerless motion analysis system for the estimation of body joint kinematics with application to sport analysis. In *10th International Conference on Computer Graphics Theory and Applications (VISIGRAPP)*, pages 58–69, 2015. 1

[10] D. Frost and J. Cronin. Stepping back to improve sprint performance: a kinetic analysis of the first step forwards. *Journal of Strength and Conditioning Research*, 25:2721–2728, 2011. 1

[11] R. Harle, J. Cameron, and J. Lasenby. Foot contact detection for sprint training. In *Asian Conference on Computer Vision 2010 (ACCV2010)*, ACCV'10, pages 297–306, Berlin, Heidelberg, 2010. Springer-Verlag. 1, 8

[12] J. G. Hay, J. A. Miller, and R. W. Canterna. The techniques of elite male long jumpers. *Journal of Biomechanics*, 10:855–866, 1986. 1

[13] S.-U. Jung and M. S. Nixon. Heel strike detection based on human walking movement for surveillance analysis. *Pattern Recogn. Lett.*, 34(8):895–902, June 2013. 2, 5

[14] S. Khan and M. Shah. Tracking multiple occluding people by localizing on multiple scene planes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):505–519, March 2009. 3

[15] M. Kim and D. Lee. Development of an imu-based foot-ground contact detection (fgcd) algorithm. *Ergonomics*, 60(3):384–403, 2017. PMID: 27068742. 1

[16] M. Kristan, A. Leonardis, J. Matas, ..., and Z. Chi. The visual object tracking vot2016 challenge results. In *Computer Vision – ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II*, pages 777–823, Cham, 2016. Springer International Publishing. 6

[17] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov 2004. 6

[18] R. Nagahara, T. Matsubayashi, A. Matsuo, and K. Zushi. Kinematics of transition during human accelerated sprinting. *Biology Open*, 3:689–699, 2014. 1

[19] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965. 5

[20] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. *Bundle Adjustment — A Modern Synthesis*, pages 298–372. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000. 3

[21] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millennium*, pages 239–269. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003. 7

[22] A. Yildiz and Y. S. Akgul. A fast method for tracking people with multiple cameras. In K. N. Kutulakos, editor, *Trends and Topics in Computer Vision: ECCV 2010 Workshops, Heraklion, Crete, Greece, September 10-11, 2010, Revised Selected Papers, Part I*, pages 128–138, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. 4

[23] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, Nov 2000. 3

[24] W. Zhu, B. Anderson, S. Zhu, and Y. Wang. A computer vision-based system for stride length estimation using a mobile phone camera. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '16, pages 121–130, New York, NY, USA, 2016. ACM. 1